Comparative genomics of banana and rice genomes: insight into an RGA cluster

Franc-Christophe BAURENS[1], Robert Neil Gerard MILLER[2], Magali LESCOT[3], Mathieu ROUARD[4], Georgios Joannis PAPPAS Jr[2], Takashi MATSUMOTO[5], Nicolas ROUX[4], Takuji SASAKI[5], Stéphanie SIDIBE BOCS[1] and Angélique D'HONT[1].

[1] UMR DAP, CIRAD, Montpellier, France
[2] Bioinformatics, UCB, Brasilia, Brazil
[3] LGPD CNRS, Marseille FRANCE
[4] CfL, Bioversity, Montpellier, France
[5] RGP, NIAS, Japan

Corresponding author : franc-christophe.baurens@cirad.fr

Introduction

Cultivated bananas are highly sterile, parthenocarpic, vegetatively propagated plants. Most of the economically important clones are polyploid and very suceptible to disease (*e.g.* fungi, bacteria, viruses).

MARGA08 probe [1], a resistance gene analog (RGA) of coiled coil - nucleic binding site - leucine rich repeat (CC-NBS-LRR) type was used to screen three banana BAC libraries.

Two from wild diploid species potentially source of resistance to Black sigatoka disease: *Musa acuminata* spp. *burmannicoides* Calcutta 4 (AA-MA4), *Musa balbisiana* "Pisang Klutuk Wulung" (BB-MbP) and one from a very susceptible triploid cultivated clone Cavendish - Grande Naine (AAA-MAC).

Positive BAC clones were subjected to BAC Fingerprint and RFLP analysis [1]. These revealed that MARGA08 is organized in cluster in the three banana clones. One BAC containing MARGA08 cluster per banana accession was sequenced and compared.

Results

The 3 BAC clones belong to the same RGA locus:

Sequence comparison using ACT and mLagan Vista [2] have shown that the 3 BACs share more than 30Kb of genomic sequence. Using automatic predictions (Eugene [3]) and subsequent manual annotation, presence of mitochondrial transcription termination factor gene (mTERF) in the 3 BACs have been reported together with presence of a Serine threonine-protein kinase gene (STK) in two of the BACs (see purple lines in Figure 4).

RGA sequence alignment:

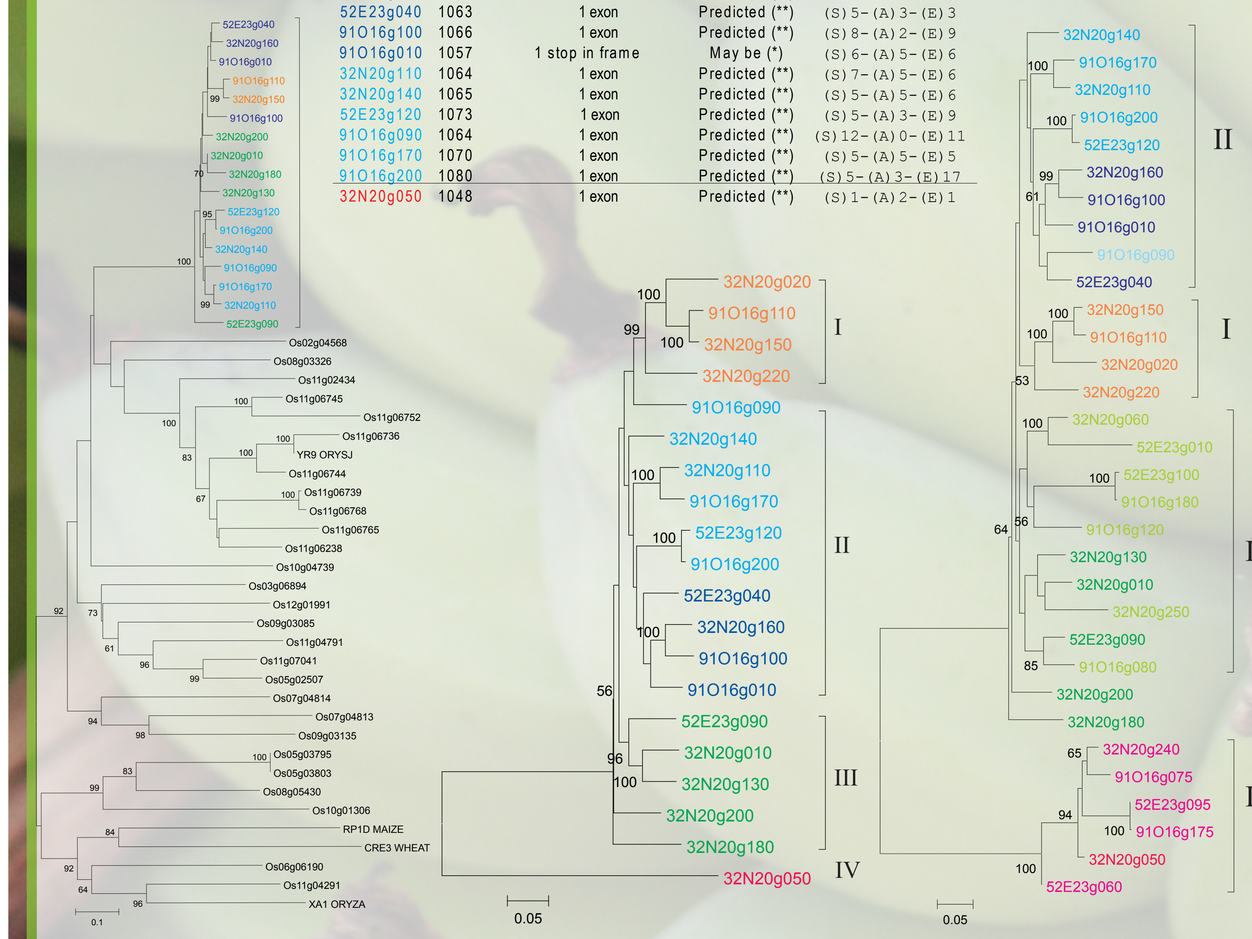
Multiple alignments were performed on the 34 RGAs using Mafft, ClustalW and finally checked manually. MARGA08 cluster (90% identity) contains both complete and fragmented RGAs in two orientations with most of the genes on the reverse strand (see Figure 4). MARGA08 coding sequences (CDS) contain several microsatellite motifs which could contribute to diversity of sequence and potentially to functionnal shifts.

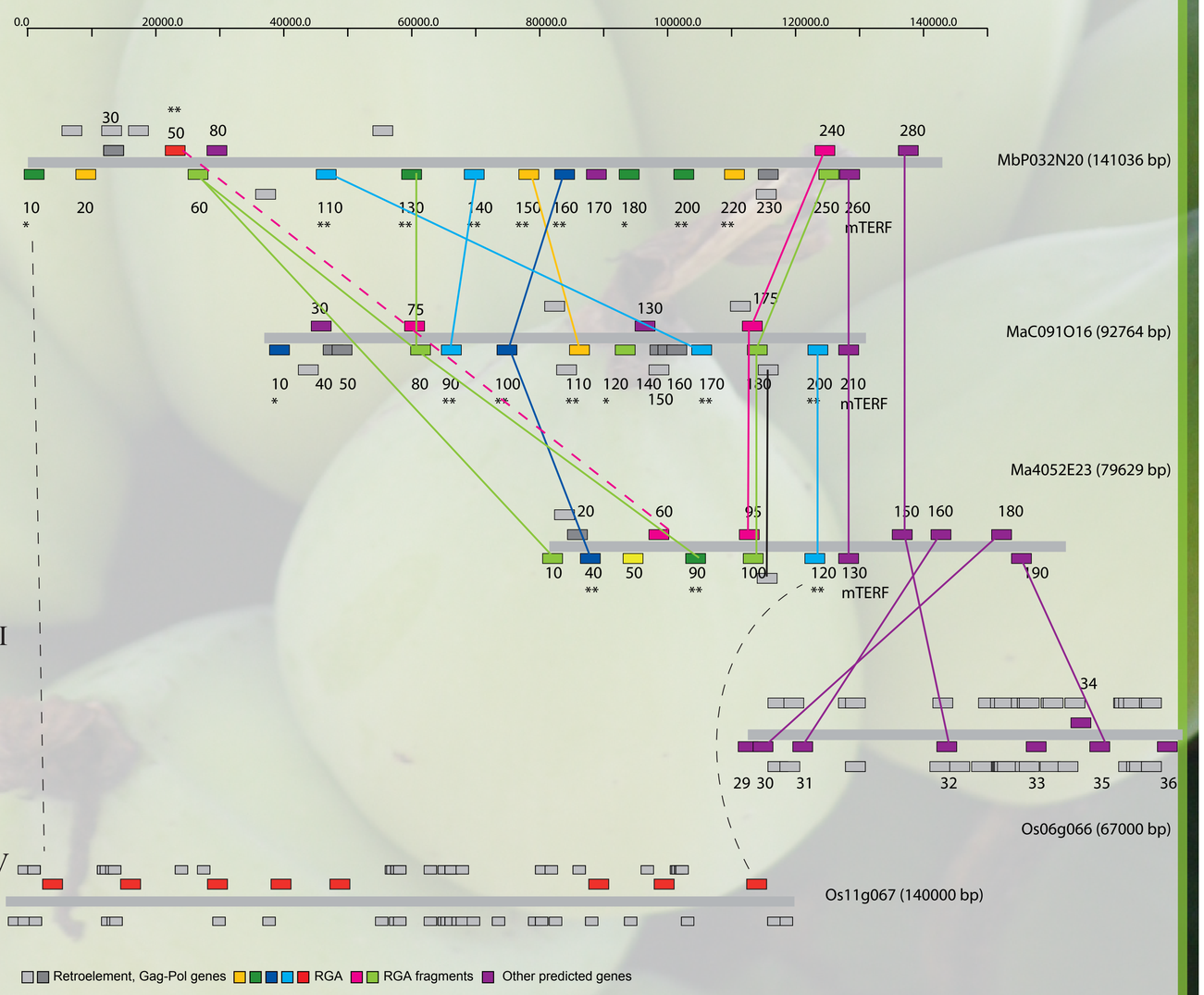
Phylogenetic analysis:

Based on sequence alignments, phylogenetic analyses have been performed using PhyML [5]. Phylogenetic analysis of both banana and rice RGAs shows that CC-NBS-LRR clusters evolved independently in the two species (see tree 1), moreover, distances within *Musa* are smaller than in rice indicating more recent duplication events in *Musa*. Four groups of genes have been defined (highlighted in color: green, blue, orange and red). 32N20g50 gene present on the direct strand is clearly separated from other RGAs present on the reverse strand (see tree 2). Analysis of MARGA08 family members including pseudogenes and fragments indicate that group IV contains all gene and fragments oriented on the direct strand (see tree 3) and group III contain all other fragments (light green).

Table Musa Resistance Gene Analog features: MARGA08 cluster [1]
L: protein length, Δ: deletion, Ψ: pseudogene

MARGA08	L (aa)	CDS structure	Expressed	Microsatellites
32N20g010	1000	2 introns, 4 small internal Δ	May be (*)	(S)14-(A)1-(E)21
32N20g130	1070	1 exon	Predicted (**)	(S)5-(A)3-(E)11
32N20g180	926	1intron, 1 large internal Δ	May be (*)	(S)3-(A)4-(E)9
32N20g200	1061	1 exon	Predicted (**)	(S)4-(A)4-(E)5
52E23g090	1057	1 exon	Predicted (**)	(S)3-(A)0-(E)4
52E23g010	793	1 retro, N-term C-term Δ	Ψ	(S)5-(A)0-(E)0
32N20g060	931	3 stops in frame, N-term Δ	Ψ	(S)9-(A)0-(E)0
91O16g120	860	N-term Δ	May be (*)	(S)6-(A)4-(E)18
32N20g020	1100	1 retro	Ψ	(S)5-(A)5-(E)4
32N20g150	1067	1 exon	Predicted (**)	(S)5-(A)4-(E)9
32N20g220	1070	1 exon	Predicted (**)	(S)12-(A)6-(E)5
91O16g110	1067	1 exon	Predicted (**)	(S)10-(A)4-(E)5
32N20g160	1072	1 exon	Predicted (**)	(S)6-(A)3-(E)4
52E23g040	1063	1 exon	Predicted (**)	(S)5-(A)3-(E)3
91O16g100	1066	1 exon	Predicted (**)	(S)8-(A)2-(E)9
91O16g010	1057	1 stop in frame	May be (*)	(S)6-(A)5-(E)6
32N20g110	1064	1 exon	Predicted (**)	(S)7-(A)5-(E)6
32N20g140	1065	1 exon	Predicted (**)	(S)5-(A)5-(E)6
52E23g120	1073	1 exon	Predicted (**)	(S)5-(A)3-(E)9
91O16g090	1064	1 exon	Predicted (**)	(S)12-(A)0-(E)11
91O16g170	1070	1 exon	Predicted (**)	(S)5-(A)5-(E)5
91O16g200	1080	1 exon	Predicted (**)	(S)5-(A)3-(E)17
32N20g050	1048	1 exon	Predicted (**)	(S)1-(A)2-(E)1





1 Phylogeny of Musa and rice NBS domain of RGAs
protein MAFFT [4], PhyML [5].

2 Phylogeny of Musa MARGA08 CC-NBS-LRR
Complete CDS manual alignment, phylogeny by PhyML [5].

3 Phylogeny of Musa MARGA08
CDS and fragment manual alignment, PhyML [5].

4 Musa - Rice RGA cluster comparison using GenomePixelizer [6]
Othologous relationships identified using BBMH, ACT, Lagan/Vista [2] and phylogenetic analysis, are indicated by colored lanes. Star indicates potential expression (see Table); no star: pseudogenes and fragments.

Conclusion and propects

The MARGA08 gene ancestor seems to be mono-exonic.

RGA cluster seems to be the result of a complex pattern of amplification involving gene duplication before and after *M. acuminata* / *M. balbisiana* speciation.

Further investigations will involve (i) the fine study of microsatellite polymorphism inside coding sequence, (ii) the determination of the type of selection affecting MARGA 08 genes by computing ω = dN/dS ratio, (iii) the complete sequencing of MARGA08 locus in *Musa*, including haplotypes and (iv) comparison with recent RGAs discovered in banana [7-8]

References:

[1] Miller et al. (submitted 2007)
[2] Sidibe Bocs et al. (ISMB congress, Vienna 2007)
[3] Foissac et al. BMC Bioinformatics 6, 25 (2005)
[4] Katoh et al. Nucleic Acids Res. 33, 511 (2005)
[5] Guindon & Gascuel. Syst Biol 52,696 (2003)
[6] Kozik et al. Bioinformatics 18, 335 (Feb 2002)
[7] Pei et al. Plant Sci. 172, 1166 (2007)
[8] Pereza-Echeverria et al. Mol. Gen. Genomics (2007)